```
import pandas as pd
import seaborn as sns
```

```
df=pd.read_csv("insurance.csv")
df
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...	...	...	...	...	...	...	...
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

1338 rows × 7 columns

Next steps: Generate code with df | New interactive sheet

```
df['age'].mean()
df
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...	...	...	...	...	...	...	...
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

1338 rows × 7 columns

Next steps: Generate code with `df` New interactive sheet

```
df.isnull().sum()
```

	0
age	0
sex	0
bmi	0
children	0
smoker	0
region	0
charges	0

dtype: int64

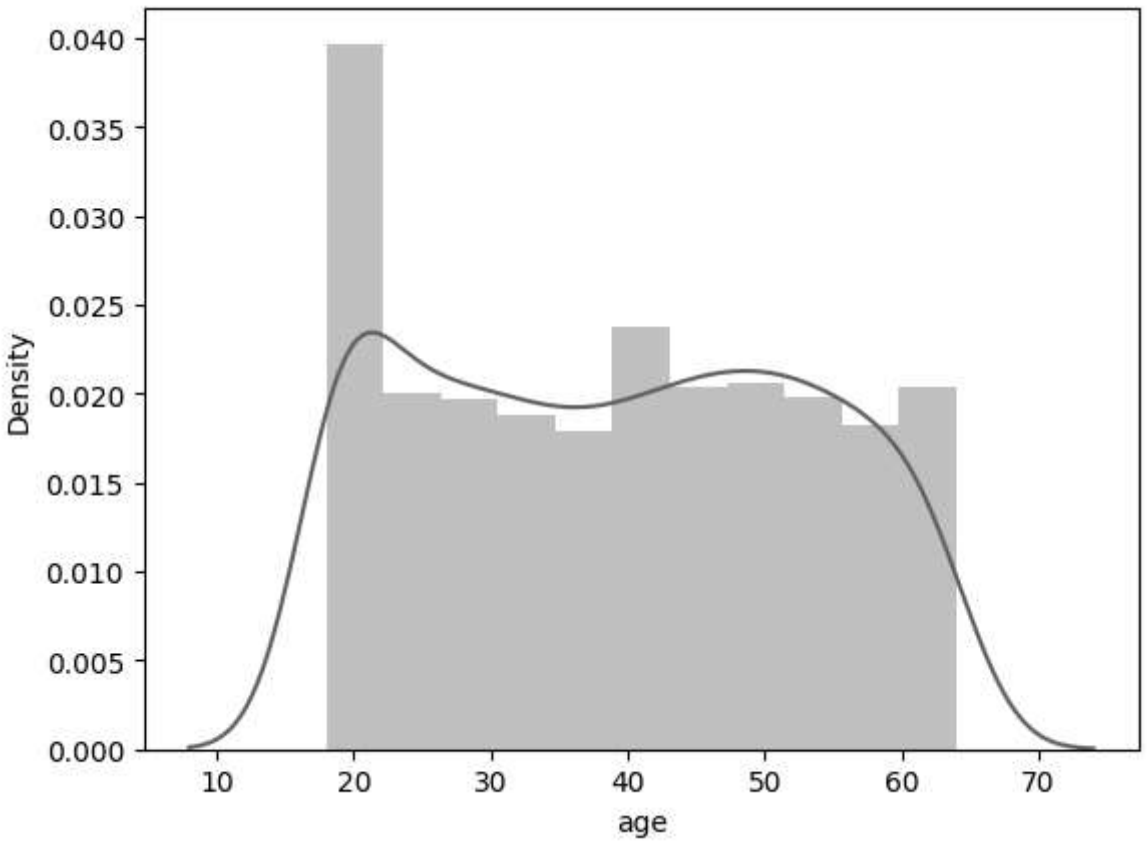
```
df['sex'].value_counts()
```

sex	count
male	676
female	662

dtype: int64

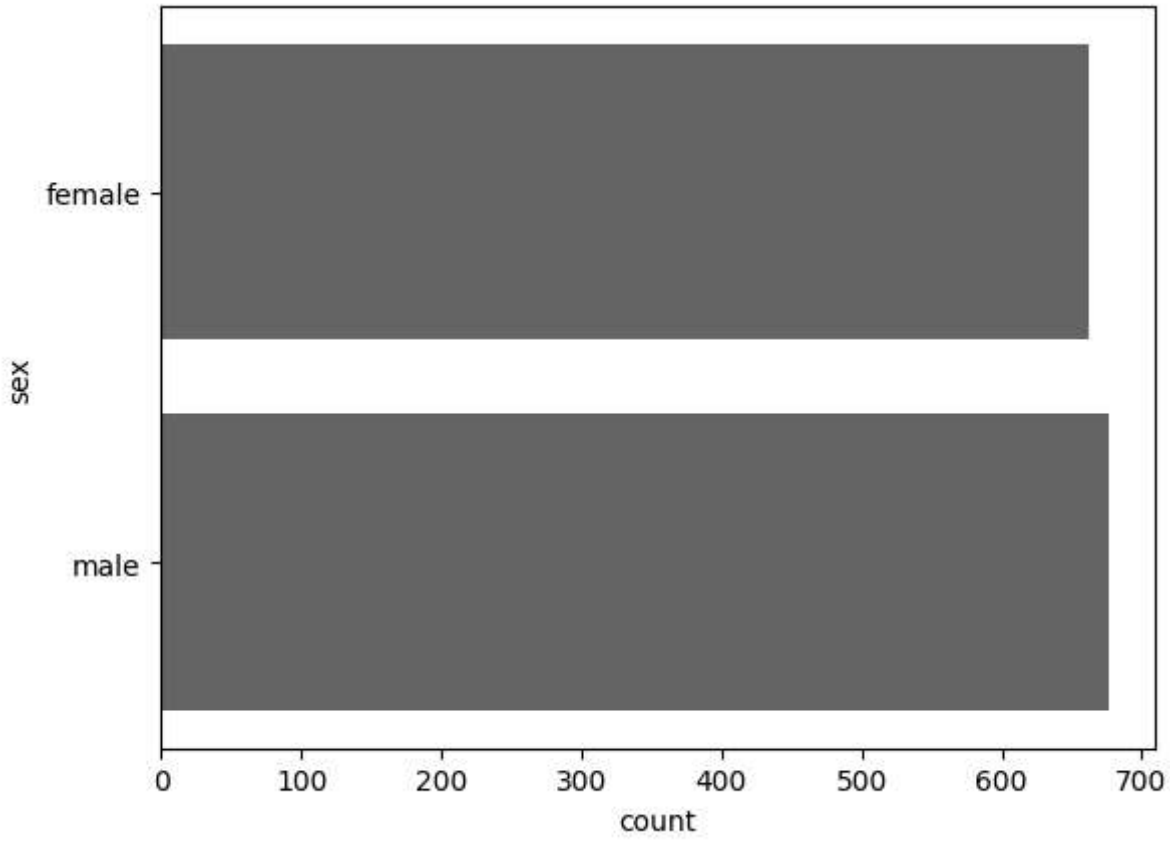
```
sns.distplot(df['age'])
```

```
/tmp/ipython-input-3234920688.py:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['age'])
<Axes: xlabel='age', ylabel='Density'>
```



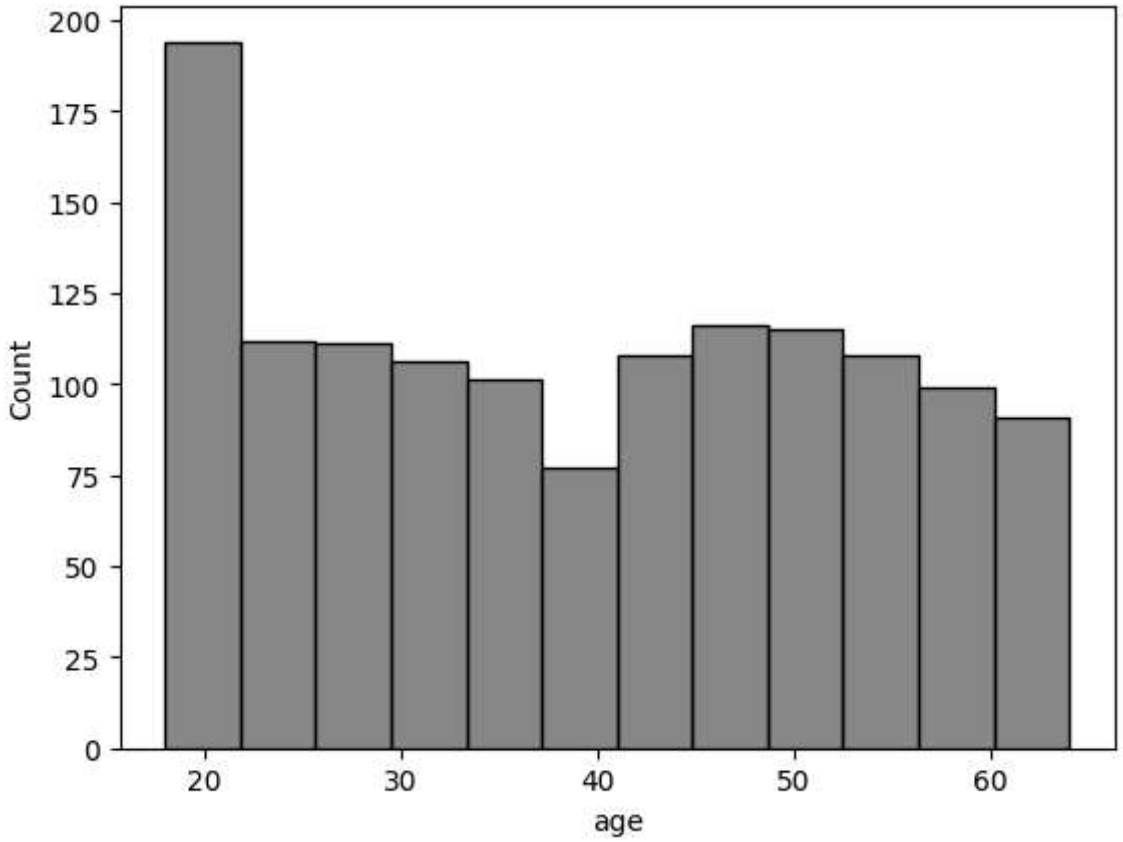

```
sns.countplot(df['sex.'])
```

```
<Axes: xlabel='count', ylabel='sex'>
```



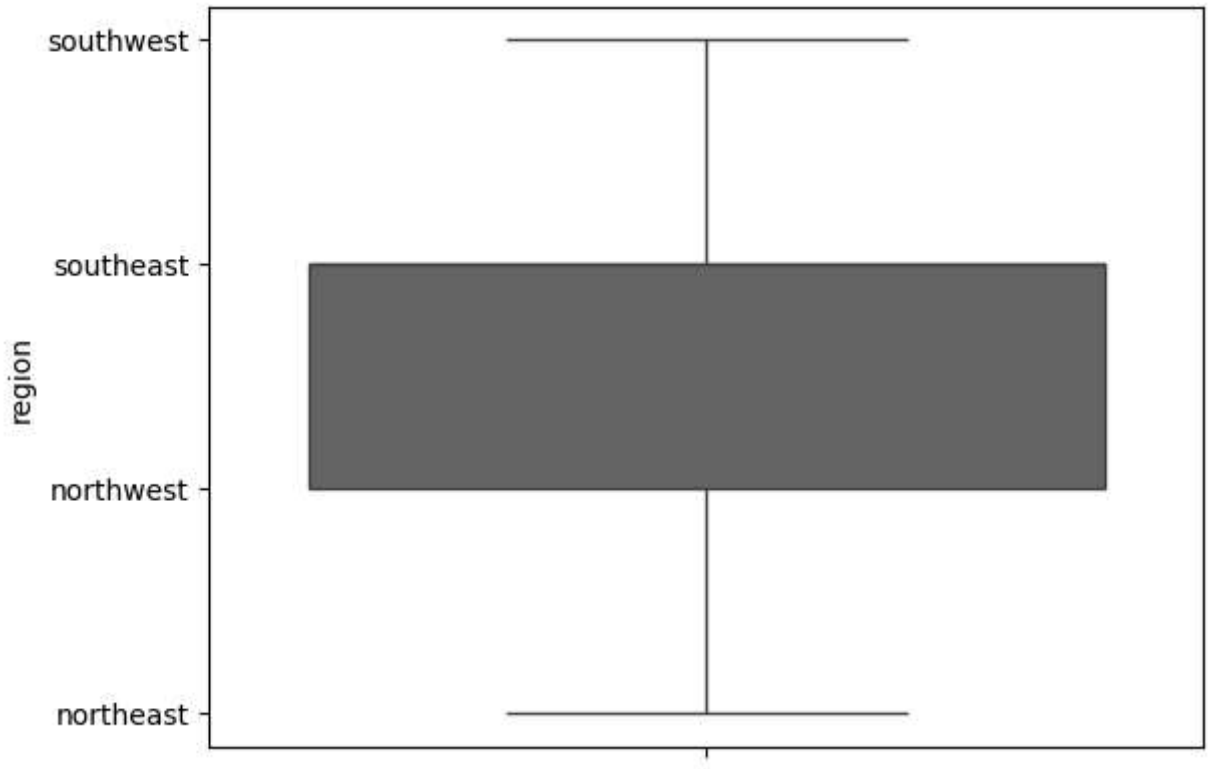

```
sns.histplot(df['age'])
```

```
<Axes: xlabel='age', ylabel='Count'>
```



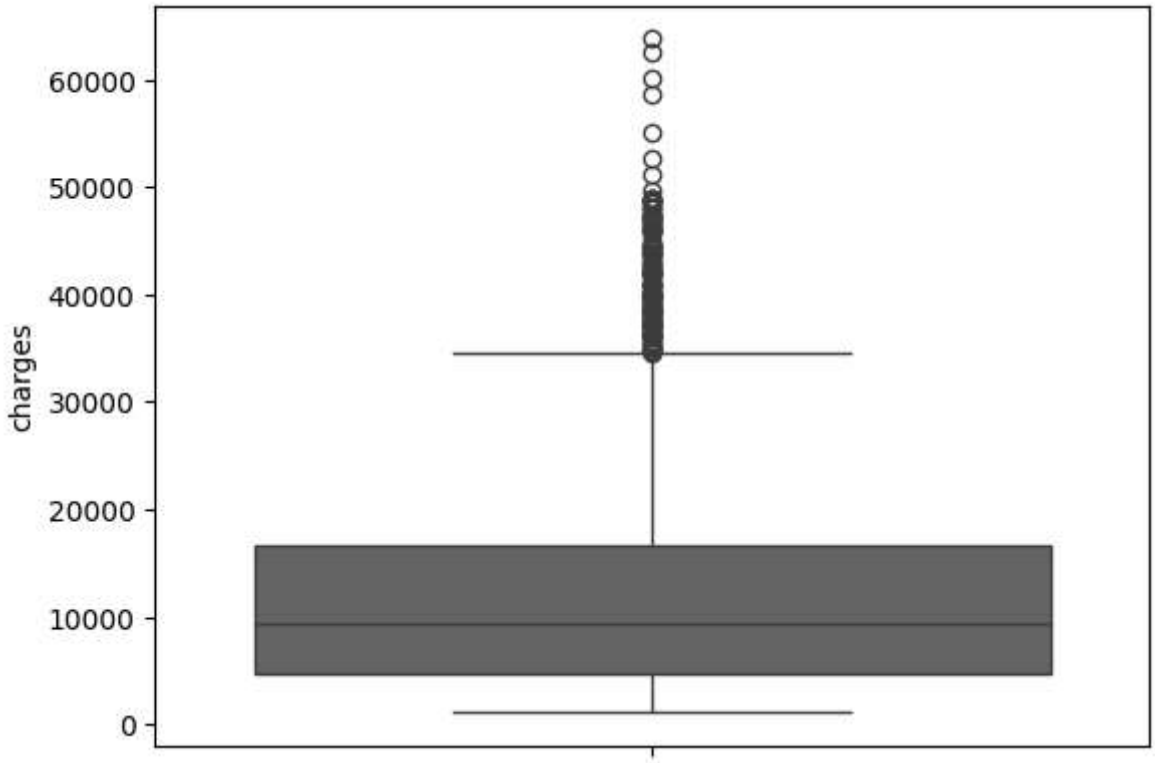

```
sns.boxplot(df['region'])
```

```
<Axes: ylabel='region'>
```



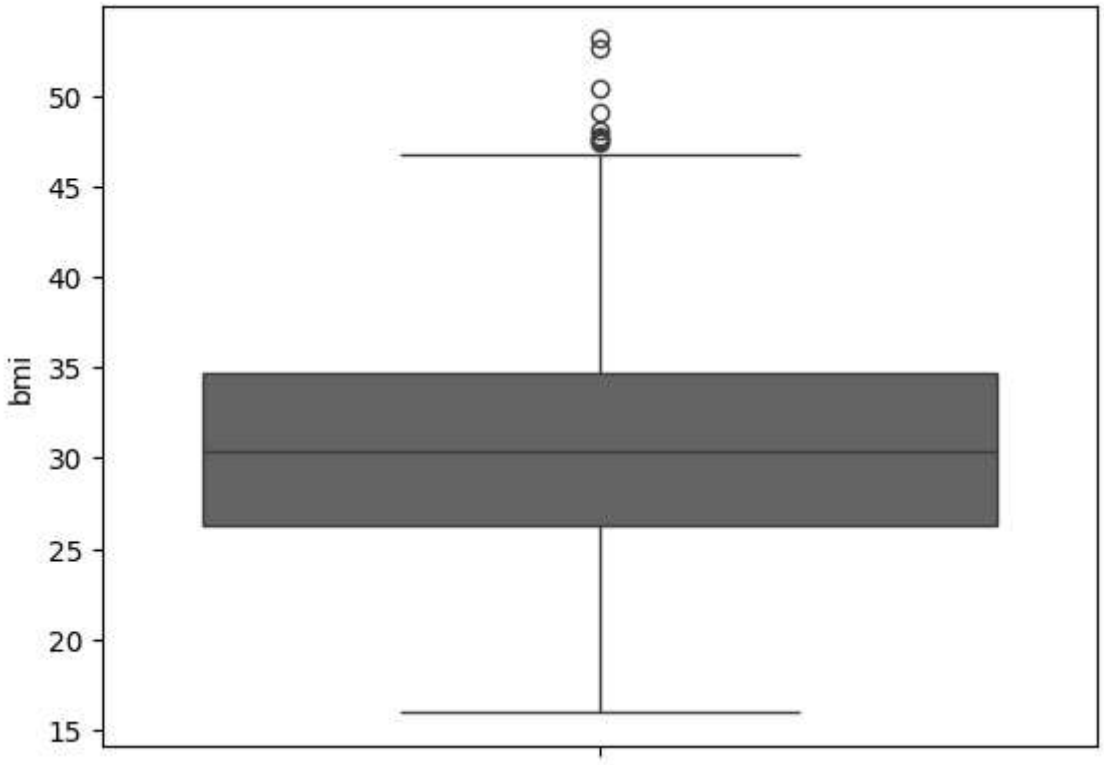
```
sns.boxplot(df['charges'])
```

```
<Axes: ylabel='charges'>
```




```
sns.boxplot(df['bmi'])
```

```
<Axes: ylabel='bmi'>
```



```
Start coding or generate with AI.
```

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df['sex']=le.fit_transform(df['sex'])
df['region']=le.fit_transform(df['region'])
df['smoker']=le.fit_transform(df['smoker'])
```

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
a=["sex","smoker","region"]
for i in a:
    df[i]=le.fit_transform(df[i])
```

```
a=[1,2,3,4,5,67,]
```

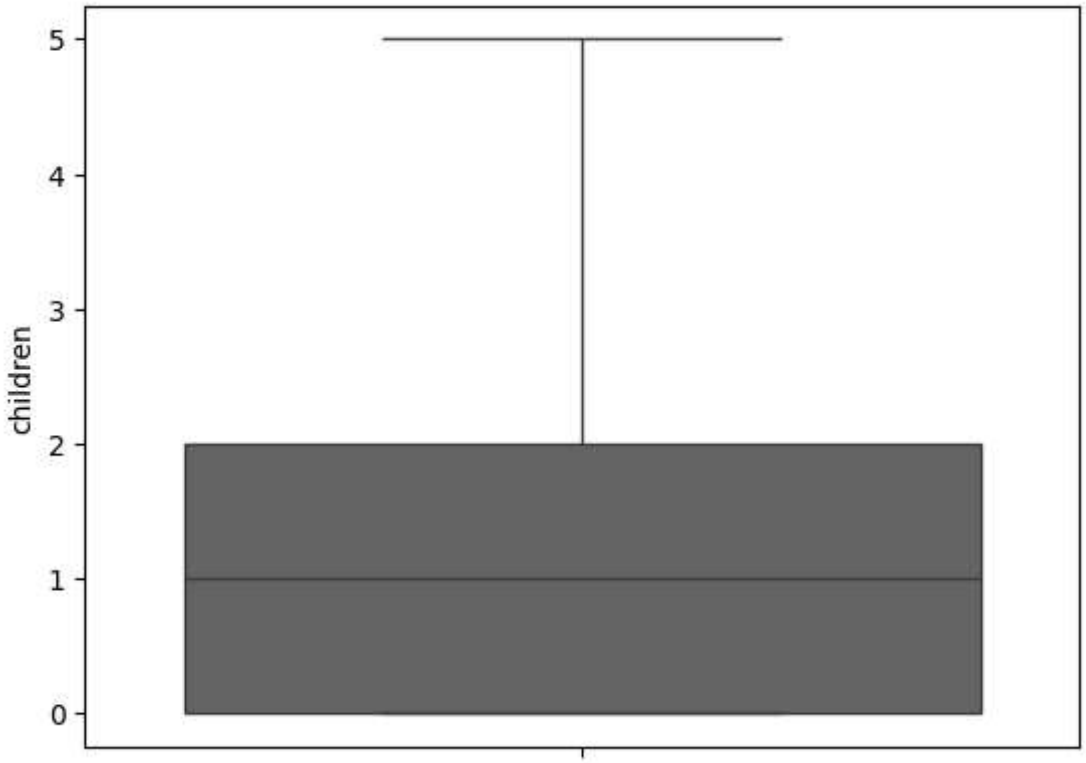
```
df.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	0	27.900	0	yes	southwest	16884.92400
1	18	1	33.770	1	no	southeast	1725.55230
2	28	1	33.000	3	no	southeast	4449.46200
3	33	1	22.705	0	no	northwest	21984.47061
4	32	1	28.880	0	no	northwest	3866.85520

Next steps: Generate code with df | New interactive sheet

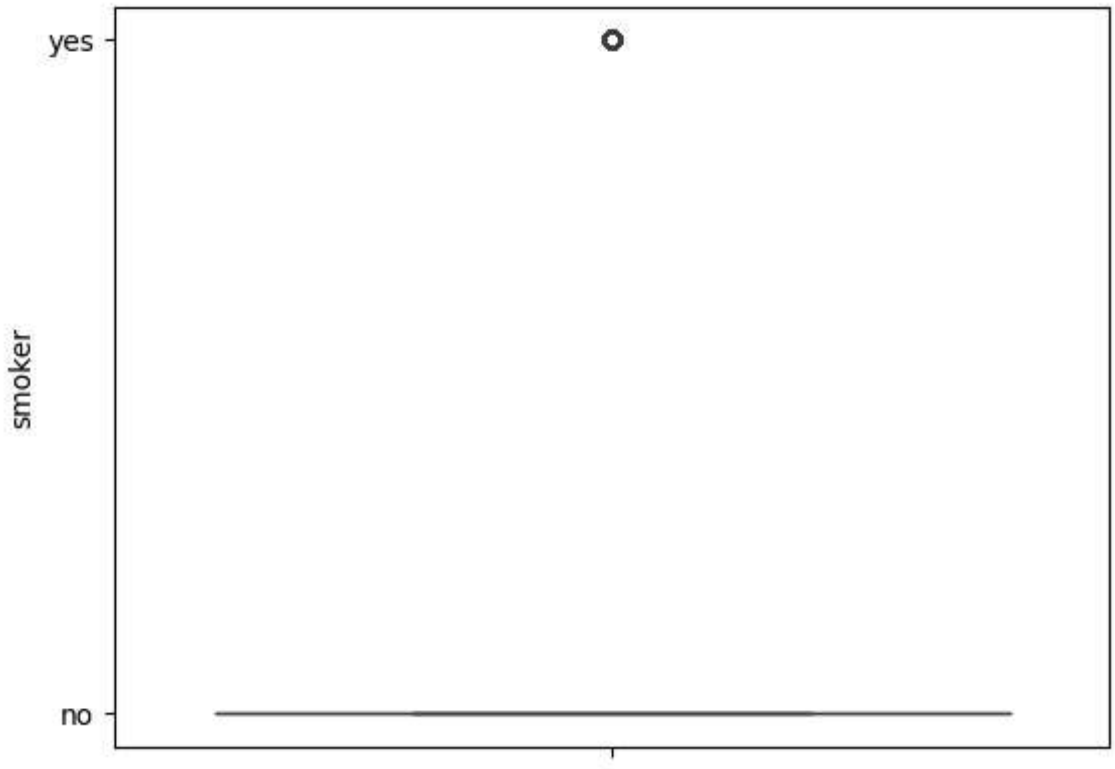
```
sns.boxplot(df['children'])
```

```
<Axes: ylabel='children'>
```




```
sns.boxplot(df['smoker'])
```

```
<Axes: ylabel='smoker'>
```




```
sns.boxplot(df['sex'])
```

```
<Axes: ylabel='sex'>
```